

Ročník 7, Číslo I., duben 2012

# ANALÝZA POUŽITELNOSTI METÓDY X-12-ARIMA PRI PROGNÓZOVANÍ A DEKOMPOZÍCII ČASOVÝCH RADOV DOPYTU CESTUJÚCICH

## ANALYSIS OF APPLICABILITY OF X-12-ARIMA METHOD FOR FORECASTING AND DECOMPOSITION OF PASSENGER DEMAND TIME SERIES

Ondrej Cyprich<sup>1</sup>, Lukáš Holeša<sup>2</sup>

*Anotácia: Identifikácia a popis trendu, sezónnosti, prebiehajúcich zmien v charaktere hlavných zložiek ekonomických časových radov ako i ich prognózovanie vyžaduje použitie vhodnej dekompozičnej metódy na ich modelovanie a prognózovanie. Hlavným cieľom príspevku je posúdiť použiteľnosť metódy X-12-ARIMA, ktorá by mala eliminovať negatívne vplyvy známych nedostatkov metódy X-11-ARIMA/88, na dekompozíciu časového radu dopytu cestujúcich po prímestskej autobusovej doprave. Použiteľnosť procedúry X-12-ARIMA bola posudzovaná sériou štatistických testov, súhrnnými mierami (M1-M7, Q1 a Q2) a mierami presnosti predikcie. Výstupy analýzy potvrdili použiteľnosť metódy X-12-ARIMA na dekompozíciu ako i prognózovanie časových radov dopytu cestujúcich po prímestskej doprave.*

*Kľúčové slová: X-12-ARIMA, Dopyt cestujúcich po prímestskej autobusovej doprave, Krátkodobé prognózovanie dopytu cestujúcich, Sezónna dekompozícia dopytu cestujúcich.*

*Summary: Identification and description of trend, seasonality, ongoing changes in nature of main components of economic time series as well as its forecasting requires application of proper decomposition approach to its modelling and forecasting. The main purpose of the paper is to analyse applicability of the X-12-ARIMA decomposition procedure, which should eliminate limitations of X-11-ARIMA/88 method, in order to decompose of passenger demand for suburb bus transport time series. Applicability of X-12-ARIMA procedure was tested by several statistical tests, summary measures (M1-M7, Q1 and Q2) and goodness-of-fit measures. Outputs of the analyses proved applicability of X-12-ARIMA procedure in the process of passenger demand time series analysis.*

*Key words: X-12-ARIMA, Passenger demand for suburb bus transport, Short-term passenger demand forecasting, Passenger demand time series seasonal adjustment.*

<sup>1</sup> Ing. Ondrej Cyprich, Žilinská univerzita v Žiline, Fakulta prevádzky a ekonomiky dopravy a spojov, Katedra cestnej a mestskej dopravy, Univerzitná 1, 010 26 Žilina, Tel.: +421415133523,  
E-mail: [ondrej.cyprich@fpedas.uniza.sk](mailto:ondrej.cyprich@fpedas.uniza.sk)

<sup>2</sup> Ing. Lukáš Holeša, Žilinská univerzita v Žiline, Fakulta prevádzky a ekonomiky dopravy a spojov, Katedra cestnej a mestskej dopravy, Univerzitná 1, 010 26 Žilina, tel. : +421415133523,  
E-mail: [lukas.holesa@fpedas.uniza.sk](mailto:lukas.holesa@fpedas.uniza.sk)

## ÚVOD

V poslednom desaťročí dochádza v Slovenskej republike k výrazným zmenám vo zvyklostiach verejnosti týkajúcich sa využívania jednotlivých foriem prímestskej dopravy. Celoštátne sčítania dopravy (1, 2, 3) ako i štatistické ročenky (4) dlhodobo naznačujú prebiehajúce zmeny trendu vývoja ukazovateľa počtu cestujúcich prepravovaných verejnou osobnou ako i individuálnou dopravou. K zmenám však nedochádza len v prípade charakteru trendu vývoja, ktorý je v čase premenlivý, ale aj v charaktere sezónnosti či iných nepravidelností dopytu cestujúcich. Náležitý popis prebiehajúcich zmien vo vývoji dopytu cestujúcich ako i ich predikcia je možná len za predpokladu aplikácie vhodnej dekompozičnej metódy napr. z triedy *X-11*. K dnešnému dňu je problematika voľby vhodnej dekompozičnej metódy z tejto triedy rozpracovaná v štúdiu (5) pričom pri dekompozícii bola použitá metóda *X-11-ARIMA/88* (6), ktorej použiteľnosť nebola potvrdená. Metóda *X-12-ARIMA* (7, 8) by vzhľadom na jej špecifiká týkajúce sa predovšetkým možnosti modelovania kalendárnych variácií mala umožniť deskripciu i predikciu prebiehajúcich zmien vo vývoji *trendovo-cyklickej zložky (TC)*, *medzimesačných variácií (MMV)* a nepravidelností dopytu cestujúcich. Cieľom príspevku je testovanie vhodnosti aplikácie dekompozičnej metódy *X-12-ARIMA* pre potreby dekompozície časových radov dopytu cestujúcich po prímestskej autobusovej doprave s mesačnou periodicitou a následne vytvorenia vhodných podkladov pre analýzu prebiehajúcich zmien v *TC* a *MMV* časových radov dopytu cestujúcich po prímestskej autobusovej doprave v podmienkach Žilinského kraja. Sekundárnym cieľom príspevku je zhodnotenie presnosti predikcie aplikáciou príslušnej dekompozičnej metódy.

## 1. ŠPECIFIKÁCIA PROBLÉMOV

Nech homogénny časový rad dopytu cestujúcich je náhodný nezáporný proces so spojitými stavmi  $Q_t$  a platí:

$$Q_t \in (0, \infty) \quad (1)$$

v diskretnom čase  $T$ , tzn. že  $T$  tvorí diskkrétne reálne hodnoty kde:

$$T = \mathbb{N} = \{t_0, \dots, t_n\} = \{0, 1, 2, 3, \dots\} \quad (2)$$

Vychádzajúc zo vzťahov 1 a 2 pod označením homogénny časový rad rozumieme postupnosť hodnôt  $\{Q_1, \dots, Q_n\}$  zo spojitého intervalu na reálnej priamke modelovaných v pravidelne chronologicky usporiadaných diskretných časových okamihoch.

Spomenuté vlastnosti  $\{Q_t; 1 \leq t \leq n\}$  vyplývajú z charakteru evidencie dopravných podnikov, ktorá býva archivovaná predovšetkým vo forme intervalových časových radov. Hodnoty  $Q_t$  sú získavané *agregáciou súčtom* (vzťah 3) čiastkových počtov cestujúcich prepravených v priebehu *časových intervalov*  $I_k; I_k \in (t-1, t)$  - najčastejšie kalendárnych mesiacov.

$$Q_t = \sum_{q \in Q_t} q; q \subset Q_t, Q_t = \{q_1, \dots, q_m\} \quad (3)$$

Kde: (popis na nasledujúcej strane)

$Q_t$  - počet cestujúcich agregovaný súčtom v priebehu  $I_k; I_k \in (t-1, t)$ ,  $t = 1, \dots, n$ ,  $k = 1, \dots, n-1$ ,

$q$  - počet cestujúcich prepravených v  $j$ -tom časovom okamžiku  $j = 1, \dots, m$ .

Kľúčovými problémami dekompozície sú podobne ako u väčšiny ekonomických časových radov (ČR) aj v prípade ČR dopytu cestujúcich kalendárne variácie a v čase premenlivý charakter trendovo-cyklickej zložky ako i medzimesačných variácií.

### 1.1 Trendovo-cyklická zložka

Premenlivý charakter trendovo-cyklickej zložky znemožňuje jej spoľahlivý odhad aplikáciou jednoduchých trendových kriviek s nemennými parametrami (priamka (vzťah 4), polynóm  $k$ -tého radu (vzťah 5)). Na druhej strane podľa Cipru (9) možno v krátkych časových úsekoch ČR aplikovať na vyrovnanie elementárne matematické krivky (tzv. lokálne trendy) avšak za predpokladu rozdielnosti parametrov v týchto úsekoch.

$$\beta_0 + \beta_1 \cdot \tau, \quad \tau = 1, \dots, n \quad (4)$$

$$\beta_0 + \beta_1 \cdot \tau + \dots + \beta_k \cdot \tau^k, \quad \tau = 1, \dots, n \quad (5)$$

Uvedený problém teda možno odstrániť aplikáciou rôznych druhov kľzavých priemerov, ktoré sa v každom pozorovaní adaptujú okamžitému priebehu časového radu.

### 1.2 Medzimesačné variácie

Je zrejmé, že agregáciou (vzťah 3) vytvorený časový rad  $n$  mesačných hodnôt dopytu cestujúcich nespĺňa podmienku (vzťah 6) rovnakej dĺžky intervalov  $I_k$ , v priebehu ktorých hodnoty  $q$  boli agregované.

$$I_k = \text{konšt.}; \quad k = 1, \dots, n-1 \quad (6)$$

Nerovnaká dĺžka kalendárnych mesiacov (31, 30, 29 a 28 dní) však nie je jediným problémom vyvolaným *MMV*. *MMV* sú podľa (10) systematickou zložkou tvoriacou pôvodný časový rad a z ich systematickosti vyplývajú možnosti odhadu a následne aj eliminácie aplikáciou širokej palety metód dekompozície časových radov. *MMV* sú vyvolané kombináciou *sezónnych (SV)* a *kalendárnych (KV) variácií*  $Q_t$ .

### 1.3 Kalendárne variácie

V praxi bývajú často nedostatočne zohľadňované *KV* dopytu vyvolané nerovnakou početnosťou jednotlivých dní (po-ne) v kalendárnych mesiacoch. Kalendárne variácie vyplývajú z časovo premenlivej kompozície kalendára a do značnej miery ovplyvňujú porovnateľnosť hodnôt skúmaného ukazovateľa v časovo príslušných obdobiach (napr. januárové hodnoty). V príspevku vychádzame z Youngom (11) vypracovanej štúdie, kde bolo z hľadiska početnosti zastúpenia jednotlivých dní v kalendárnych mesiacoch identifikovaných 22 typov mesiacov (tab. 1) opakujúcich sa v 28-ročných cykloch (v súčasnosti prebiehajúci cyklus – 1/2000-12/2027).

Tab. 1 - Absolútna početnosť kalendárnych dní v 22 základných typoch mesiacov

počty príslušných dní v mesiaci									počty príslušných dní v mesiaci								
k. m.	p. d. m.	po	ut	st	št	pi	so	ne	k. m.	p. d. m.	po	ut	st	št	pi	so	ne
31-dňové mesiace									30-dňové mesiace								
1	po	5	5	5	4	4	4	4	8	po	5	5	4	4	4	4	4
2	ut	4	5	5	5	4	4	4	9	ut	4	5	5	4	4	4	4
3	st	4	4	5	5	5	4	4	10	st	4	4	5	5	4	4	4
4	št	4	4	4	5	5	5	4	11	št	4	4	4	5	5	4	4
5	pi	4	4	4	4	5	5	5	12	pi	4	4	4	4	5	5	4
6	so	5	4	4	4	4	5	5	13	so	4	4	4	4	4	5	5
7	ne	5	5	4	4	4	4	5	14	ne	5	4	4	4	4	4	5
februáre priestupných rokov									februáre nepriestupných rokov								
15	po	5	4	4	4	4	4	4	22	po-ne	4	4	4	4	4	4	4
16	ut	4	5	4	4	4	4	4									
17	st	4	4	5	4	4	4	4									
18	št	4	4	4	5	4	4	4									
19	pi	4	4	4	4	5	4	4									
20	so	4	4	4	4	4	5	4									
21	ne	4	4	4	4	4	4	5									

Vysvetlivky:

k. m. - kód mesiaca

p. d. m. - prvý deň mesiaca

Zdroj: Spracované autormi podľa Younga (11)

Analýzou kalendára boli v tab. 2 k mesiacom (jan.-dec.) rokov (2000-2012) podľa ich kompozície z hľadiska počtu dní priradené kódy mesiacov (1-22).

Tab. 2 - Typy mesiacov z hľadiska početností zastúpenia jednotlivých dní

rok	jan	feb	mar	apr	máj	jún	júl	aug	sep	okt	nov	dec
2000	6	16	3	13	1	11	6	2	12	7	10	5
2001	1	22	4	14	2	12	7	3	13	1	11	6
2002	2	22	5	8	3	13	1	4	14	2	12	7
2003	3	22	6	9	4	14	2	5	8	3	13	1
2004	4	21	1	11	6	9	4	7	10	5	8	3
2005	6	22	2	12	7	10	5	1	11	6	9	4
2006	7	22	3	13	1	11	6	2	12	7	10	5
2007	1	22	4	14	2	12	7	3	13	1	11	6
2008	2	19	6	9	4	14	2	5	8	3	13	1
2009	4	22	7	10	5	8	3	6	9	4	14	2
2010	5	22	1	11	6	9	4	7	10	5	8	3
2011	6	22	2	12	7	10	5	1	11	6	9	4
2012	7	17	4	14	2	12	7	3	13	1	11	6

Zdroj: Autori

Pozn. Typy mesiacov sú označené kódom, použitím ktorého v tab. 1 možno priradiť počet jednotlivých dní v príslušnom mesiaci.

#### 1.4 Sezónne variácie

SV sú vyvolané na jednej strane nerovnakou dĺžkou kalendárnych mesiacov ako i samotnou pravidelne opakujúcou sa nerovnomernosťou  $Q_t$  v čase. Podobne ako v prípade TC zložky aj hodnoty *sezónnych faktorov*  $S_{ij}$  mesiacov  $j$  ( $j = 1, \dots, 12$ ) nie sú v priebehu rokov  $i$  ( $i = 1, \dots, r$ ) konštantné. Zmeny hodnôt sezónnych indexov sú vyvolané zmenou návykov cestujúcich v čase a v nadväznosti nato aj zmenou ponuky zo strany dopravcu. Zohľadnenie

tohto faktu vzhľadom na preukázaný (12) viac ako 65 % podiel  $SV$  na celkovej variabilite ČR  $Q_t$  je nevyhnutné.

Špecifickým typom  $MMV$  sú školské prázdniny, tie na jednej strane za predpokladu rovnakej časovej polohy v priebehu roka možno pripísať na vrub  $SV$ . Na strane druhej, prázdniny s každoročne premenlivou časovou polohou (jarné, veľkonočné, chrípkové a pod.) alternujú medzi rôznymi časovými úsekmi v rámci rokov. Tieto je vhodnejšie pripísať na vrub  $KV$ . Konkrétne v prípade veľkonočných prázdnin medzi marcom, aprílom resp. na rozmedzí marca a apríla. Vzhľadom na rozdielnú ponuku prepravných služieb v priebehu prázdnin je potrebné ich vplyv zohľadniť nielen pri ukazovateli cestujúcich prepravených za žiacke cestovné ale aj u ostatných skupín cestujúcich, ktoré sú rozdielnou ponukou aj keď nepriamo ovplyvnené. Nezanedbateľný vplyv na vývoj dopytu majú aj tzv. sprievodné javy spojené s dňami pracovného pokoja alebo školskými prázdninami, pred začatím ktorých obyvatelia, pracujúci, resp. študujúci v iných regiónoch Slovenska zvyknú hromadne cestovať domov k svojim blízkym (Vianoce, Sviatok všetkých svätých a pod.).

Dostatočné zohľadnenie spomínanej trojice problémov ako i iných by malo byť dosiahnuteľné, za predpokladu minimalizácie ostatných vplyvov, aplikáciou dekompozičných metód triedy  $X-11$ . V tomto príspevku sú prezentované výstupy testovania použiteľnosti metódy  $X-12-ARIMA$ .

## 2. POUŽITÝ MATERIÁL A METÓDY

V tejto kapitole sú v stručnosti popísané dátové vstupy dekompozičnej úlohy vrátane vykonaných úprav, metód testovania vlastností skúmaných ČR ako i testovania použiteľnosti zvolenej dekompozičnej techniky.

### 2.1 Vstupné údaje dekompozície

Vstupné údaje analýzy boli časopriestorovo vymedzené údaje o počtoch prepravených cestujúcich ako i druhu cestovného, za ktoré boli prepravení. Tieto dáta boli evidované a archivované zariadeniami na výdaj cestovných lístkov spolupracujúceho dopravcu zabezpečujúceho prímestskú autobusovú dopravu v rámci Žilinského samosprávneho kraja. Tieto hodnoty boli neskôr dopravcom agregované súčtom tak, aby výstupom procesu agregácie boli ČR  $\{Q_p(t); 1 \leq t \leq 60\}$  mesačných hodnôt dopytu príslušných kategórií cestujúcich (obdobie mesiacov 1/2006 – 12/2010). Vhodnosť aplikácie metódy  $X-12-ARIMA$  bola teda posudzovaná na vzorke 24 ČR dopytu cestujúcich prepravených za obyčajné (občianske), žiacke a iné cestovné ako i celkových objemov dopytu cestujúcich (v každej kategórii cestovného 6 ČR). Pre zjednodušenie v ďalšom texte budeme na označenie ČR dopytu cestujúcich, pokiaľ nebude uvedené ináč, používať funkčný zápis  $Q_p(t)$ , kde  $p = 1, 2, \dots, 24$ .

Hodnoty takto vytvorených  $Q_p(t)$  boli považované za priestorovo a vecne homogénne, nakoľko sa v priebehu špecifikovaného obdobia nemenila geografická pôsobnosť dopravcu ani technológia prepravy cestujúcich v rozsahu ovplyvňujúcom jeho homogenitu z vecného a priestorového aspektu. Ďalšie úpravy  $Q_p(t)$  pre zabezpečenie homogenity, preto neboli nutné. Spomedzi analyzovanej vzorky  $Q_p(t)$  štvorica ČR obsahovala chýbajúcu

hodnotu pre mesiac jún 2006. Problém chýbajúcich hodnôt bol vyriešený zaradením parametra (parametra aditívnych odľahlých hodnôt) do *regARIMA* modelu príslušného ČR a následne nahradením chýbajúceho pozorovania hodnotou dostatočne veľkou, aby bolo považované za odľahlé v procese odhadu parametrov *regARIMA* modelu. Po odhade parametrov *regARIMA* modelu boli v rámci procedúry *X-12-ARIMA* chýbajúce hodnoty nahradené ich odhadmi. Pre potreby predikcie ako i dekompozície procedúra ďalej pracovala s pôvodnými ako i odhadnutými hodnotami skúmaných ČR.

## 2.2 Použité metódy

Sledujúc splnenie cieľa výskumnej aktivity boli v závislosti od vlastností skúmaných časových radov použité *multiplikatívne ako i aditívne varianty dekompozičnej metódy X-12-ARIMA* (7, 8). Dekompozícia ako i testy uvedené v tejto podkapitole boli realizované v programovom vybavení SAS 9.1.3 (13).

### 2.2.1 Nastavenia procedúry X-12-ARIMA

Všeobecné špecifikácie procedúry *X-12-ARIMA* sú zrejmé z nasledujúceho programového skriptu:

```
ods graphics on;
title height=3 "X12 ARIMA-obycajne_NO-Testovanie pouzitelnosti sezonneho
ocistovania";
proc x12 data=sasuser.XY date=date interval=month seasons=12 notrimmiss;
var XY;
transform power=0;
regression predefined=(constant td seasonal);
automdl maxorder=(4,2) maxdiff=(2,1) print=(UNITROOTTEST AUTOCHOICE
UNITROOTTESTMDL AUTOCHOICEMDL BEST5MODEL)ljungboxlimit=0.95 armacv=0.15;
estimate;
outlier type=(AO LS TC) CV=3.63;
forecast lead=12;
x11 seasonalma=msr outforecast final=all mode=mult;
output out=out A1 A2 A6 A8 A8AO A8LS A8TC B1 C17 C20 D1 D7 D8 D9 D10 D10D
D11 D12 D13 D16 D16B D18 E5 E6 E7 MV1;
run;
```

Voľbu aditívneho resp. multiplikatívneho variantu procedúry *X-12-ARIMA* sme vykonali v závislosti od výstupov *heuristických testov potreby logaritmickkej transformácie* (14, 15) jednotlivých  $Q_p(t)$ . V prípadoch použitia multiplikatívneho variantu (`x11 seasonalma=msr outforecast final=all mode=mult`) bola použitá logaritmická transformácia (`transform power=0`) pôvodných hodnôt časových radov. Pri aplikácii aditívneho variantu (`x11 seasonalma=msr outforecast final=all mode=add`) nebola použitá žiadna transformácia (`transform power=1`). Parametre základných komponentov *regARIMA* modelu (`regression predefined=(constant td seasonal)`) boli ponechávané pri potvrdení ich štatistickej významnosti na hladine významnosti  $\alpha = 0,05$ , v opačnom prípade boli nevýznamné zložky z modelu odstraňované. Maximálny rád nesezónnych a sezónnych *ARMA* polynómov v procese automatickej identifikácie *SARIMA* modelu bol v prípade chybových hlásení procedúry znižovaný pri nesezónnych (*AR*, *MA*) zo 4 na 2



a sezónnych (*SAR*, *SMA*) z 2 na 1. Rovnako sme postupovali aj v prípade špecifikácie maximálneho rádu diferencovania hodnôt  $Q_p(t)$ , kde sme znižovali rád nesezónnej diferencie (*I*) z 2 na 1 a sezónnej diferencie (*SI*) z 1 na 0. Pôvodný zápis v skripte `automdl maxorder=(4,2) maxdiff=(2,1)` sme teda pri úpravách maximálneho rádu polynómov a diferencií pri niektorých  $Q_p(t)$  upravili do podoby `automdl maxorder=(2,1) maxdiff=(1,0)`. V niektorých prípadoch sme na základe odporúčaní chybových hlásení procedúry pristúpili k zvýšeniu počtu iterácií z 200 na 1500 (`estimate maxiter=1500`). Kritická hodnota (*CV*) pre identifikáciu odľahlých (extrémnych) hodnôt bola určená podľa *Ljungovej* (16) rovnice vzhľadom na počet pozorovaní analyzovaných  $Q_p(t)$   $n = 60$  na  $CV = 3,63$ . Ostatné nastavenia procedúry boli pri výpočtoch nemenné. Ďalšie podrobnosti o jednotlivých zložkách navrhnutého skriptu sú podrobne rozpracované v návode použitého programového vybavenia (15). Kompletne skripty pre jednotlivé  $Q_p(t)$  sú súčasťou dizertačnej práce (17).

Vzhľadom na nedostupnosť dát o sezónnom vzore resp. charakteristikách nerovnomernosti  $Q_p(t)$  v priebehu dní kalendárneho týždňa boli regresné koeficienty modelu popisujúceho rozdielne zloženie čo sa počtu jednotlivých dní v kalendárnych mesiacoch týka (*trading day regression models (TDRegr)*), odhadnuté procedúrou automaticky. Štatistická významnosť modelov *regARIMA* bola testovaná aplikáciou *F*-štatistik *F*-testu štatistickej významnosti modelu (17) a významnosť parametrov týchto modelov vyhodnocovaná testami štatistickej významnosti regresných koeficientov  $\beta_j$  ( $\beta_j = 1, 2, \dots, 7$ ) na 5% hladine významnosti. Odhad parametrov modelov *TDRegr* bol vykonaný aplikáciou *podmienenej metódy najmenších štvorcov (CLS)*.

### 2.2.2 Nastavenia testov

Multiplikatívnosť resp. aditívnosť vývoja  $Q_p(t)$  boli preukazované aplikáciou subjektívnych (*analýza grafu*) ako i objektívnych (*heuristické testy potreby logaritmické transformácie* (14, 15)) metód.

Existencia a charakter sezónnosti boli testované aplikáciou *F*-testu na prítomnosť stabilnej (*FTSS*) a premenlivej sezónnosti (*FTPS*) (18) a kombinovaného testu na prítomnosť identifikovateľnej sezónnosti (*KTIS*) (19) (kombinácia predošlých dvoch testov a *Kruskalového-Wallisového testu na prítomnosť stabilnej sezónnosti*). Testy boli vykonané na 5% hladine významnosti.

Vhodnosť aplikácie dekompozičnej metódy *X-12-ARIMA* bola posudzovaná analýzou grafov spektier ( $G1$ ,  $G2$ ) (21):

- diferencovaných, transformovaných sezónne očistených hodnôt  $Q_p(t)$ ,
- a modifikovaných hodnôt iregulárnej zložky  $Q_p(t)$ .

Analýzou spektier bola posudzovaná prítomnosť *SV* ako i *KV*. V zmysle odporúčaní popísaných Hoodom (21) bola dekompozícia považovaná za nevyhovujúcu, ak aspoň jedna z hodnôt spektra sezónnych ako i kalendárnych variácií dosahovala špičkové hodnoty. Prítomnosť špičiek bola posudzovaná v prípade *SV* pre  $k/12$  cyklov ( $1 \leq k \leq 5$ ) za mesiac a pri

*KV* hodnoty 0,348; 0,432 cyklov za mesiac podľa Clevelanda (22) a 5,2 cyklov za mesiac podľa výstupov procedúry *X-12-ARIMA* použitého programového vybavenia.

Ďalej bola vhodnosť aplikácie metódy *X-12-ARIMA* posudzovaná tradičnými diagnostikami modelu (*ACF* (autokorelačná funkcia) reziduí analyzovaných ČR, *t-testy* významnosti parametrov navrhnutých modelov) a aplikáciou *M1-M7*, *Q1* a *Q2 súhrnných štatistík* popísaných Lothianom a Morrym (23). Realizácia *SSA* (Sliding spans analysis (24)) nebola vykonaná vzhľadom na nedostatočný počet pozorovaní ČR, ktoré boli predmetom dekompozície. Pričom dekompozíciu procedúra na základe hodnôt *M* a *Q štatistík* automaticky označovala za vyhovujúcu, čiastočne vyhovujúcu, čiastočne nevyhovujúcu alebo nevyhovujúcu.

### 2.2.3 Vyhodnotenie presnosti predikcie

Presnosť predikcie bola posudzovaná komparáciou mediánových hodnôt priemernej absolútnej percentuálnej chyby (*MAPE*, vzťah 7) prognostických výstupov procedúry *X-12-ARIMA* získaných priamou dekompozíciou (ČR  $Q_p(t)$  prepravených za obyčajné, žiacke, iné cestovné ako i ich súčet za celú dopravnú spoločnosť) a nepriamou dekompozíciou (ČR  $Q_p(t)$  prepravených za obyčajné, žiacke, iné cestovné ako i ich súčet za jednotlivé prevádzky dopravcu) s prognostickými výstupmi najpresnejšieho modelu identifikovaného automatizovaným prognostickým nástrojom *TSFS* (Time Series Forecasting System) programového vybavenia SAS 9.1.3 s rozšírenou modelovou základňou (*SARIMA* modely, modely exponenciálneho vyrovnávania, regresné modely, celkom 242 modelov). Podrobné špecifikácie a nastavenia automatického prognostického nástroja sú súčasťou dizertačnej práce autora (17).

$$MAPE = \frac{1}{h} \cdot \sum_{T=n+1}^{n+h} \frac{|Q_T - \hat{Q}_T|}{Q_T} \cdot 100 \quad [\%] \quad (7)$$

Kde:

$Q_T$  - skutočná hodnota ČR dopytu cestujúcich pre obdobie  $T = n + h$ ;  $n = 60$ ;  $h = 1, 2, \dots, 12$ ,

$\hat{Q}_T$  - modelom odhadnutá hodnota ČR dopytu cestujúcich pre obdobie  $T = n + h$ ;  $n = 60$ ;  $h = 1, 2, \dots, 12$ ,

$h$  - horizont prognózy jednotlivých  $Q_p(t)$ ,  $h = 1, 2, \dots, 12$ ,

$n$  - počet pozorovaní  $Q_p(t)$  použitých na návrh prognostických modelov,  $n = 60$ .

## 3. VÝSTUPY RIEŠENIA A DISKUSIA

Heuristickými testami bola potvrdená prítomnosť sezónnej a v niektorých prípadoch aj inej ako konštantnej trendovo-cyklickej zložky (viď. tab. 3). Prítomnosť týchto vlastností podporuje hypotézu, že sezónna dekompozícia aplikáciou metódy *X-12-ARIMA* by mohla mať určitý význam aj v prípade analyzovaných  $Q_p(t)$ . Výstupy heuristických testov taktiež preukázali prevládajúci multiplikatívny charakter (vzťah 8b) analyzovaných  $Q_p(t)$ . Výnimku tvorili niektoré ČR dopytu cestujúcich prepravených za iné cestovné, ktoré sa vyznačovali aditívnosťou vzájomného vzťahu ich *TC*, *SV*, *KV* a *I* zložky (vzťah 8a). V závislosti od identifikovaných vlastností sme pri ďalších výpočtoch volili multiplikatívne resp. aditívne varianty metódy *X-12-ARIMA*.



$$a) Q_p(t) = TC_t + D_t + S_t + I_t \quad b) Q_p(t) = TC_t \cdot D_t \cdot S_t \cdot I_t \quad (8)$$

Tab. 3 – Identifikované vlastnosti analyzovanej vzorky časových radov dopytu cestujúcich 2006-2010

časový rad vlastnosť	žiacke_A	obyčajné_A	iné_A	spolu_A	žiacke_B	obyčajné_B	iné_B	spolu_B
M/A	M	M	A	M	M	M	A	M
trend	-	+	-	-	-	+	-	+
sezónnosť	+	+	+	+	+	+	+	+
časový rad vlastnosť	žiacke_C	obyčajné_C	iné_C	spolu_C	žiacke_D	obyčajné_D	iné_D	spolu_D
M/A	M	M	M	M	M	M	M	M
trend	-	+	+	-	-	+	-	+
sezónnosť	+	+	+	+	+	+	+	+
časový rad vlastnosť	žiacke_E	obyčajné_E	iné_E	spolu_E	Σ_žiacke	Σ_obyčajné	Σ_iné	Σ_spolu
M/A	M	M	A	M	M	M	M	M
trend	-	+	-	+	-	+	-	+
sezónnosť	+	+	+	+	+	+	+	+

Zdroj: Výstupy výpočtov autorov

Pozn. M - multiplikatívny, A – aditívny charakter časového radu  $Q_p(t)$ ; trend resp. sezónnosť „-“ – nie je prítomný(á), „+“ – je prítomný(á).

Výstupy testov *štatistickej významnosti parametrov regARIMA modelu* (tab. 4) poukázali na skutočnosť, že v ani jednom prípade nemožno všetky parametre (v tab. 4 označené - *konštanty* (C), *sezónnych* (SV) a *kalendárnych variácií* (KV)) *regARIMA* modelov  $Q_p(t)$  považovať za štatisticky významné. Navyše štatistická významnosť parametrov kalendárnych variácií odhadnutých procedúrou *X-12-ARIMA* bola vyhodnotená vo väčšine prípadov ako nevyhovujúca. Za jednoznačne štatisticky významné možno považovať len odhady parametrov konštanty a odľahlých hodnôt *regARIMA* modelov  $Q_p(t)$ . Spomedzi dvanástich parametrov sezónnych variácií až na výnimky neboli vo vzorke vyhodnocovaných *regARIMA* modelov  $Q_p(t)$  modely s viac ako dvomi parametrami vyhodnotenými ako štatisticky nevýznamnými. Skupinové testy parametrov sezónnych variácií (*SkSV*) a kalendárnych variácií (*SkTD*) poukázali aj napriek čiastočnej štatistickej nevýznamnosti parametrov SV a KV na významnosť zaradenia príslušných zložiek do *regARIMA* modelov  $Q_p(t)$ . Výstupy testov sezónnosti (*FTSS/FTPS/KTIS* tab. 4), konkrétne *F-testu na prítomnosť stabilnej* (*FTSS*) a *premenlivej* (*FTPS*) *sezónnosti* a *kombinovaného testu na prítomnosť identifikovateľnej sezónnosti* (*KTIS*) potvrdzujú prítomnosť stabilnej sezónnosti pri všetkých analyzovaných ukazovateľoch. Potvrdenie stabilnej sezónnosti by malo garantovať spoľahlivý odhad sezónnych faktorov. Sezónnosť v reziduálnej zložke modelov nebola potvrdená. Hodnoty štatistík *M1-M7*, *Q1* a *Q2* drvicej väčšiny *regARIMA* modelov  $Q_p(t)$  nasvedčovali v prospech použiteľnosti metódy *X-12-ARIMA* pri riešení dekompozície ČR dopytu. Grafy spektier kalendárnych variácií poukázali pri niektorých *regARIMA* modeloch  $Q_p(t)$  na

nedostatky, nakoľko bola pri ich analýze potvrdená prítomnosť špičiek najmä pri hodnotách 3,6 a zriedkavejšie 4,2 a 5,2 cyklov za mesiac (stĺpce NF-G1, NF-G2 tab. 4).

Tab. 4 – Výstupy testov použiteľnosti *X-12-ARIMA*

časový rad \ test	regARIMA				SKSV	SKTD	ARIMA	FTSS	FTPS	KTIS	RS	Q1	Q2	NŠ M1-M7	NF-G1	NF-G2
	C	SV	TD	OUT												
žiacke_A	+	+	-/+	+	+	+	+	+	+	+	+	+/-	+/-	2	5,2	+
obyčajné_A	+	+/-	-/+	o	+	+	+/-	+	+	+	+	+	+	1	4,2	4,2
iné_A	o	+/-	-/+	o	+	+	+	+	+	+	+	+/-	-/+	2	3,6	3,6
spolu_A	+	o	-/+	+	o	+	+	+	+	+	+	+	+	1	5,2	5,2
žiacke_B	+	+/-	-/+	+	+	+	+	+	+	+	+	+	+/-	2	+	+
obyčajné_B	+	+/-	-/+	o	+	+	+	+	+	+	+	+	+	+	+	+
iné_B	o	+/-	-/+	+	+	+	+	+	+	+	+	+	+	+	+	+
spolu_B	+	+/-	-/+	o	+	+	+	+	+	+	+	+	+	+	+	+
žiacke_C	o	+	-/+	+	+	+	+/-	+	+	+	+	+	+	2	+	+
obyčajné_C	+	+/-	-/+	o	+	+	+/-	+	+	+	+	+	+	+	+	+/-
iné_C	+	+/-	+/-	+	+	+	+	+	+	+	+	+	+	+	+	+
spolu_C	+	+/-	-/+	+	+	+	+/-	+	+	+	+	+	+	2	+	+
žiacke_D	+	o	-/+	o	+	+	-	+	+	+	+	+	+	2	+	+
obyčajné_D	+	+/-	-/+	o	+	+	+/-	+	+	+	+	+	+	+	3,6	3,6
iné_D	o	+/-	+/-	+	+	+	+	+	+	+	+	+	+	+	+	+
spolu_D	o	o	-/+	o	o	+	+	+	+	+	+	+	+	+	+	+
žiacke_E	+	+	-/+	+	+	+	+	+	+	+	+	+/-	+/-	+	+	+
obyčajné_E	+	o	-/+	o	o	+	+	+	+	+	+	+	+	+	+	+
iné_E	o	+/-	-/+	o	+	+	+	+	+	+	+	+	+	+	+	+
spolu_E	+	+/-	-/+	o	+	+	+	+	+	+	+	+	+	1	+	+
Σ_žiacke	+	o	-/+	+	o	+	+/-	+	+	+	+	+	+/-	3	+	+
Σ_obyčajné	+	+/-	-/+	+	+	+	+/-	+	+	+	+	+	+	+	+	+
Σ_iné	o	+/-	+/-	+	+	+	+/-	+	+	+	+	-	-	5	3,6	3,6
Σ_spolu	+	o	-/+	+	o	+	+	+	+	+	+	+	+	1	+/-	+

Zdroj: Výstupy výpočtov autorov

Legenda:

- „o“ - vypracovanie procedúry vzhľadom na neprítomnosť príslušnej zložky *regARIMA* modelu nebolo nutné,
- „+“ - výstupy testov príslušnej procedúry jednoznačne nasvedčujú v prospech aplikácie dekompozície metódou *X-12-ARIMA*,
- „-“ - výstupy testov príslušnej procedúry jednoznačne nasvedčujú v neprospech aplikácie dekompozície metódou *X-12-ARIMA*,
- „+/-“ - výstupy testov príslušnej procedúry neposkytujú jednoznačné výsledky avšak viac ako polovica výstupov týchto testov nasvedčuje skôr v prospech vhodnosti aplikácie dekompozície metódou *X-12-ARIMA*,
- „-/+“ - výstupy testov príslušnej procedúry neposkytujú jednoznačné výsledky avšak viac ako polovica výstupov týchto testov nasvedčuje skôr v prospech nevhodnosti aplikácie dekompozície metódou *X-12-ARIMA*.

Z porovnania mediánových hodnôt *MAPE* (tab. 5) vypočítaných pre priamu, nepriamu dekompozíciu *X-12-ARIMA* ako i automatizovaný prognostický nástroj programového

vybavenia SAS je zrejmé, že pri prognózovaní ČR  $Q_p(t)$  je za účelom dosiahnutia presného prognostického výsledku najvýhodnejšie použiť priamu dekompozíciu  $X-12-ARIMA$  v prípadoch:

- dopytu cestujúcich prepravených za obyčajné cestovné,
- a dopytu cestujúcich prepravených za žiacke cestovné.

Naproti tomu automatizovaný prognostický nástroj programového vybavenia SAS s rozšírenou modelovou základňou je najvhodnejšie použiť v prípadoch:

- dopytu cestujúcich prepravených za iné cestovné,
- a celkového dopytu cestujúcich.

Z vypočítaných mediánových hodnôt  $MAPE$  (tab. 5) je taktiež zrejmé, že aplikácia priamej dekompozície  $X-12-ARIMA$  je vždy výhodnejšia v porovnaní s nepriamym prístupom k dekompozícii. Uvedené je spôsobené nižšími hodnotami  $MAPE$  priamej dekompozície v porovnaní s nepriamou.

Tab. 5 – Porovnanie mediánových hodnôt  $MAPE$  prognóz ex-ante jednotlivých prístupov

ukazovateľ metodológia	žiacke [%]	obyčajné [%]	iné [%]	spolu [%]
$P_{+rA}$	6,81	14,70	6,56	7,95
$L_{+rA}$	6,03	4,78	5,90	7,72
$A_{min}$	6,04	5,39	4,41	2,05

Zdroj: Výstupy výpočtov autorov

Kde:

$P_{+rA}$  - mediánové hodnoty  $MAPE$  získané nepriamou dekompozíciou  $X-12-ARIMA$ ,

$L_{+rA}$  - mediánové hodnoty  $MAPE$  získané priamou dekompozíciou  $X-12-ARIMA$ ,

$A_{min}$  - mediánové hodnoty  $MAPE$  získané automatickým prognostickým nástrojom  $TSFS$  SAS.

## ZÁVER

Výstupy výpočtov posudzujúcich vhodnosť aplikácie metódy  $X-12-ARIMA$  poukázali na existenciu nedostatkov pri jej použití na prognostické resp. dekompozičné účely. Prakticky ani pri jednom ukazovateli neboli splnené všetky požiadavky kladené na požadované štatistické vlastnosti modelov. Aj napriek uvedenej skutočnosti však priama dekompozícia metódou  $X-12-ARIMA$  poskytuje lepšie prognostické výstupy v porovnaní s automatizovaným prognostickým nástrojom  $TSFS$  programového vybavenia SAS s rozšírenou modelovou základňou. Pri aplikácii metódy  $X-12-ARIMA$  na skúmanú vzorku časových radov dopytu cestujúcich predĺžených o 12 novo dostupných pozorovaní (mesačné hodnoty ukazovateľa za rok 2011) bolo zaznamenané významné zlepšenie štatistických vlastností modelov. V prospech vhodnosti použitia tohto dekompozičného prístupu svedčia aj hodnoty súhrnných štatistík  $MI-M7$ ,  $Q1$  a  $Q2$ , ktoré s výnimkou jedného časového radu takmer bez výhrad potvrdili vhodnosť použitia tejto procedúry.

Na základe skutočností prezentovaných v príspevku bola do určitej miery potvrdená použiteľnosť dekompozičnej metódy *X-12-ARIMA* na dekompozíciu a prognózovanie časových radov dopytu cestujúcich po prímestskej autobusovej doprave v regióne Žilinského kraja. V budúcom období by mali byť výskumné aktivity zamerané na použitie tejto dekompozičnej metódy s vopred zadefinovanými parametrami kalendárnych variácií *regARIMA* modelu, čím by sa mohlo dosiahnuť ďalšie spresnenie numerických výstupov tohto postupu. Realizácia tohto výskumu však vyžaduje dostupnosť časových radov denných počtov prepravených cestujúcich jednotlivých kategórii, ktoré v súčasnosti neboli zo strany dopravcu na výskum poskytnuté. Potvrdenie vhodnosti aplikácie procedúry *X-12-ARIMA* zároveň umožnilo použitie jej výstupov na identifikáciu a deskripciu charakteru zložiek skúmaných časových radov a v nich prebiehajúcich zmien. Tieto informácie môžu byť ďalej využité pri manuálnom návrhu konštrukcie modelov časových skúmaných ukazovateľov.

## POUŽITÁ LITERATÚRA

- (1) *Celoštátne sčítanie dopravy 2010* [online]. c2010 [cit. 2011-10-20].  
Dostupné z <<http://www.ssc.sk/sk/Rozvoj-cestnej-siete/Dopravne-inzinierstvo/Celostatne-scitanie-dopravy-2010.ssc>>.
- (2) *Celoštátne sčítanie dopravy 2005* [online]. c2005 [cit. 2011-10-20].  
Dostupné z <<http://www.ssc.sk/sk/Rozvoj-cestnej-siete/Dopravne-inzinierstvo/Celostatne-scitanie-dopravy-v-roku-2005.ssc>>.
- (3) *Celoštátne sčítanie dopravy 2000* [online]. c2000 [cit. 2011-10-20].  
Dostupné z <<http://www.ssc.sk/sk/Rozvoj-cestnej-siete/Dopravne-inzinierstvo/Celostatne-scitanie-dopravy-v-roku-2000.ssc>>.
- (4) *Štatistická ročenka Slovenskej republiky* [online]. c1996-2000 [cit. 2011-10-29].  
Dostupné z <<http://portal.statistics.sk/showdoc.do?docid=15949>>.
- (5) CYPRICH, O.. *Analýza použiteľnosti metódy X-11-ARIMA/88 pri sezónnom očisťovaní dopytu cestujúcich po prímestskej autobusovej doprave*, 2001. [zatiaľ nepublikované]
- (6) DAGUM, E.B. *The X11ARIMA/88 Seasonal Adjustment Method : Foundations and User's Manual*. Ottawa : Statistics Canada, 1988.
- (7) U.S. BUREAU OF THE CENSUS. *X-12-ARIMA Reference Manual, Version 0.2.8*. Washington, DC : SAS U.S. Bureau of the Census, 2001.
- (8) U.S. BUREAU OF THE CENSUS. *X-12-ARIMA Seasonal Adjustment Program, Version 0.2.8*. Washington, DC : SAS U.S. Bureau of the Census, 2001.
- (9) CIPRA, T. *Finanční ekonometrie*. 1. vyd. Praha : Ekopress, 2008. 538 s. ISBN 978-80-86929-43-9.
- (10) TREVIN, D. A. *Guide to Interpreting Time Series – Monitoring Trends*. Canberra : ABS, 2003. 148 s. ISBN 0-642-54286-4.
- (11) YOUNG, A.. Estimating Trading-Day Variation. In *Monthly Economic Time Series*. Washington, D. C. : U.S. Bureau of the Census, 1965. Technical paper No. 12.
- (12) CYPRICH, O. *Modelovanie dopytu cestujúcich po prímestskej autobusovej doprave* [koncept dizertačnej práce]. Žilinská univerzita v Žiline. Fakulta prevádzky a ekonomiky

- dopravy a spojov; Katedra cestnej a mestskej dopravy. Školiteľ: prof. Ing. Štefan Liščák, CSc. . Stupeň odbornej kvalifikácie: philosophiae doctor. Žilina fPEDaS, 2012.
- (13) SAS 9.1.3 [počítačový program na disku]. Cary, NC : SAS Institute Inc., 2003. posledná aktualizácia 22. december 2008.
- (14) HALLAHAN, Ch. The Time Series Forecasting System In *SUGI 25*. part Hands-on Workshop, 2000. Paper 155-25.
- (15) SAS INSTITUTE Inc. *SAS/ETS® 9.1 User's Guide*. 1. vyd. Cary, NC : SAS Institute Inc., 2004. 2436 s. ISBN 1-59047-244-6.
- (16) LJUNG, G. M. On Outlier Detection in Time Series In *Journal of the Royal Statistical Society. Series B (Methodological)*, 1993. Vol. 55, No. 2. s. 559-567.
- (17) ŠOLTÉS, E. *Regresná a korelačná analýza s aplikáciami*. 1. vyd. Bratislava : Iura, 2008. 538 s. ISBN 978-80-86929-43-9.
- (18) CLEVELAND, W.S., DEVLIN, S.J. Calendar Effects in Monthly Time Series : Detection by Spectrum Analysis and Graphical Methods In *Journal of the American Statistical Association*, 1980. Vol. 75, No. 371. s. 487-496.
- (19) DAGUM, E.B. *The X-11-ARIMA Seasonal Adjustment Method*. Statistics Canada, 1980.
- (20) AKAIKE, H., ISHIGURO, M. Bayesian Seasonal Adjustment Program In *Computer Science Monographs*. Tokyo : The Institute of Statistical Mathematics, 1980. No. 13.
- (21) HOOD, C. C. SAS Programs to Get the Most from X-12-ARIMA's Modeling and Seasonal Adjustment Diagnostics In *SUGI 25*, 2000. part Statistics and Data analysis. Paper 264-25.
- (22) CLEVELAND, W.S., Devlin, S.J. Calendar Effects in Monthly Time Series : Detection by Spectrum Analysis and Graphical Methods In *Journal of the American Statistical Association*, 2000. Vol. 75, No. 371. s. 487-496.
- (23) LOTHIAN, J., MORRY M. *A Set of Quality Control Statistics for the X-11-ARIMA Seasonal Adjustment Method*. Research paper 78-10-002E. s.l. : Statistics Canada, 1978.
- (24) FINDLEY, D.F., MONSELL, B.C., SHULMAN, H.B., Pugh, M.G. Sliding Spans Diagnostics for Seasonal and Related Adjustments In *Journal of the American Statistical Association*, 1990. Vol 85, No. 410. s. 345-355.